

NOVA

IMS

Information
Management
School

MGI

Master Degree Program in
Information Management

BEYOND TURING'S TEST: DEVELOPING A STRATEGY AS A GUIDE FOR AI-HUMAN DISCRIMINATION

Maria Catarina de Almeida e Silva Mateus da Silva

Master Thesis

presented as partial requirement for obtaining the Master Degree in Information Management

NOVA Information Management School
Instituto Superior de Estatística e Gestão de Informação

Universidade Nova de Lisboa

NOVA Information Management School
Instituto Superior de Estatística e Gestão de Informação
Universidade Nova de Lisboa

**BEYOND TURING'S TEST: DEVELOPING A STRATEGY TO GUIDE FOR AI-HUMAN
DISCRIMINATION**

by

Maria Catarina de Almeida e Silva Mateus da Silva

Master Thesis presented as partial requirement for obtaining the Master's degree in
Information Management, with a specialization in Information Systems and Technologies Management

Supervised by

Vitor Duarte dos Santos

Nova IMS – Information Management School

July, 2024

STATEMENT OF INTEGRITY

I hereby declare having conducted this academic work with integrity. I confirm that I have not used plagiarism or any form of undue use of information or falsification of results along the process leading to its elaboration. I further declare that I have fully acknowledged the Rules of Conduct and Code of Honor from the NOVA Information Management School.

Lisbon, July 2024

ACKNOWLEDGEMENTS

Firstly, I would like to thank my family for their constant support and involvement throughout my academic journey. To my partner and our dog, thank you for all the support and the lightness you brought to this stage in my academic life. To my dear friends, thank you for all the motivation and support.

To the individuals who participated in the strategy execution phase, I appreciate your time and apologize for testing out your patience.

Finally, a big thank you to my thesis supervisor, Professor Vitor Santos. Thank you for your guidance, shared knowledge, motivation, and support throughout the thesis. Thank you for your time, dedication, and availability over the past year.

ABSTRACT

This thesis pursues the development of a strategy that aims to be a tool of support for everyday individuals and business owners, in their daily interactions, to identify when their subject of conversation is artificial intelligence. Artificial Intelligence has been on the rise for the last years and has become part of all of our lives, either in work, social media or even in the news. It's a topic that we might know of, but are not familiar with or comfortable around. This study delved into understanding Artificial Intelligence as a concept and focused on gathering information regarding existing tests for the discrimination of AI from humans and ethics concerns regarding the topic. As it is an evolving topic, an all-around agreeable test hasn't been developed that is able to accurately differentiate Artificial Intelligence from humans, and that task becomes even more challenging with the rapid evolution of the area. The thesis proposes a strategy made up of multiple tests that aim to support everyday individuals and businesses in their daily interactions. The strategy did not provide an absolute certain result; however, it did conclude and evaluate that testing humans and the machine on their patience with dealing with repetitive questions and disrespectful behavior, there is a noticeable response difference between both subjects.

KEYWORDS

Artificial Intelligence; Turing Test; Ethics; Human Intelligence; Human Experience; Thoughts; Emotions; AI Hallucination

Sustainable Development Goals (SDG):



TABLE OF CONTENTS

Statement of Integrity	ii
Acknowledgements	iii
Abstract	iv
List of Figures.....	vii
List of Tables.....	viii
List of Abbreviations and Acronyms.....	ix
1. Introduction.....	1
1.1. Background and problem identification.....	1
1.2. Objectives	2
1.3. Importance and Relevance.....	3
2. Literature review	4
2.1. Artificial Intelligence.....	4
2.1.1. Concepts Overview.....	4
2.1.2. Philosophy of Human Intelligence and Experience.....	6
2.2. Artificial and Human Intelligence – A Systematic Literature Review.....	6
2.2.1. PRISMA Methodology	6
2.2.2. PRISMA Execution	7
2.2.2.1. Keywords.....	7
2.2.2.2. Search String	8
2.2.2.3. Search Databases.....	8
2.2.2.4. Research Criteria.....	8
2.2.3. PRISMA Analysis	9
2.2.4. PRISMA Discussion	12
2.2.4.1. Status of the research in Human and Artificial Intelligence	12
2.2.4.2. Ethic concerns regarding Artificial Intelligence	13
2.2.4.3. Current strategies and tests to evaluate Artificial Intelligence	14
3. Methodology	17
3.1. Design Science Research	17
3.2. Research Strategy.....	18
4. Proposal of a guide for the differentiation OF AI from Humans.	20
5. Execution and evaluation	24
5.1. Use Case	24
5.1.1. Test A – Request a plan based on false premises:	24
5.1.2. Test B – Irony/Sarcasm.....	25

5.1.3. Test C – AI Hallucination.....	26
5.1.4. Test D – AI Patience.....	27
5.2. Evaluation and Discussion	33
6. Conclusions.....	35
Bibliographical References	36
Appendix A	40

LIST OF FIGURES

Figure 1 - Affective social robot, meet Sophia, an example of an AfC system.	5
Figure 2 - PRISMA 2020 Flow Diagram - BMJ 2021	7
Figure 3 - PRISMA Execution	9
Figure 4 - Example of a pair of Winograd Schema Challenge questions.	15
Figure 5 - DSR Process	18
Figure 6 - Strategy Flow Chart.....	22
Figure 7 - Demonstration of strategy execution	32

LIST OF TABLES

Table 1 – Systematic Review Research Keywords	7
Table 2 - Systematic Review Research Database	8
Table 3 - Systematic Review Research Criteria	8
Table 4 - Systematic Review Research Results.....	10
Table 5 - Evaluation of the strategy final scores	21
Table 6 - Result of each subject score	30
Table 7 – New proposed score results	34

LIST OF ABBREVIATIONS AND ACRONYMS

AI	Artificial Intelligence
LLM	Large Language Model
AfC	Affective Computing
DSR	Design Science Research
HI	Human Intelligence

1. INTRODUCTION

The introduction chapter of this study presents the background of the research topic, singling out the existing problems and unanswered questions around said topic. Then, proceeding into the identification of the study objectives and presenting the relevance of conducting the research in this thesis. Detailing the importance of understanding the objectives and context of the research before commencing the study.

1.1. BACKGROUND AND PROBLEM IDENTIFICATION

Artificial Intelligence started to be a topic of discussion since the creation of the digital computer in 1940s (Copeland, 2023). As seen in the article from Copeland, lastly updated in 2023, Alan Turing, a British mathematician and logician, that had an important role in the war against Germany deciphering the Germans encrypted messages by developing a systematic method to break the code. He was a pioneer in the development of digital computers, programming and became a founding father of Artificial Intelligence. Since then, many advances have been made in Artificial Intelligence, but only recently as this topic been accessible to anyone with a computer and internet. Large Language Model (LLM) has “revolutionized” the Industry 4.0 that we are currently in (Anadkat, 2023). ChatGPT, a LLM, is a software that responds to questions using natural language and is able to provide answers with the same conversational language as a human (Gregersen, 2023). This advance has brought an awareness to this technology and all its possible impacts in the world we know nowadays.

The appearance of ChatGPT has been controversial in many different areas. There has been a rise in websites that create art, videos that impersonate real humans, imitate voices and much more. The potential of this innovation is immense, as it is creating new jobs, aiding businesses to be able to resolve issues and speed productivity, allow students to deepen their knowledge in a subject independently, and overall enabling up to date, prompt knowledge to anyone. All this innovation has caused new potential threats and risks to the life we know and has been causing new ethical questions on whether this evolution is driving humans in the right direction.

In some news from Browns’ article, written by Kevin Stacey in 2021, awareness is raised on how AI is no longer only a field of scientific study, and it has become intertwined with our everyday lives in the more recent years. This calls for a great urgency in deciphering what potential negative impacts it might bring. As stated in the report, it is crucial for computer scientists and experts in social sciences and law to work together in order to understand, mitigate and minimize the downfalls of this new technology.

The known threats that have been identified range from hackers impersonating people we know - as an article published by The Washington Post, written by Pranshu Verma, tells the story of an elderly couple that was robbed by scammers using AI to impersonate their grandson requesting for bail money; students submitting work for evaluation written by AI; AI

generated art causing controversy due to the lack of regulations around the subject and the uncertainty it might cause to professionals, as shown in an article by Rachel Metz (2022); the spread of misinformation and potential biases when approaching decision making in judicial and health system, topics discussed in an article published by Stanford University (2021), and many others.

All these risks drive to a common need, being able to distinguish AI from Humans. It's necessary to have mechanisms and tools that allow us to distinguish "the man" vs. "the machine". Alan Turing created the Turing Test to determine whether a computer can "think". To "think" is a human capability and characteristic. The Turing test is based on a scenario where a human interrogator needs to distinguish between a human subject and a computer. The higher the probability of the computer to be confused with the human, the closer the computer is to "thinking". However, with the appearance of ChatGPT and all other LLM, this test must be reevaluated for its capability to assess a computer "thought" and possibly the need to find the "next" Turing Test of our generation.

Gathering all these thoughts, do we consider LLM software, such as ChatGPT, to have "thoughts", if not, can we confirm that the Turing Test is outdated? What about the AI hallucination, the term currently used to describe the ability for this model to answer to questions with inaccurate or misinformation, is that a sign that AI will be able to create new information to respond to an inquiry that it originally doesn't have the required stored knowledge to answer it correctly? Will there be a time where it becomes impossible to distinguish? Can we be replaced? The ultimate question to be answered, what mechanism can we create to distinguish AI from Humans?

These interrogations lead to the formulation of the following research question:

RQ: What strategies can be followed to distinguish AI from Humans?

1.2. OBJECTIVES

The goal of the research is to develop a strategy to guide the differentiation between the most advanced AI from Humans.

In order to achieve this goal, the following intermediate objectives were defined:

- Framework philosophy of AI and its main concepts;
- Make a comprehensive study on the existing strategies and test to distinguish AI from Humans;
- Create a strategy to guide distinguish AI from Humans;
- Build a use case;
- Validate the strategy.

1.3. IMPORTANCE AND RELEVANCE

Having concluded that the Turing Test, once the standard for assessing the machine's ability to show intelligence (Oppy & Dowe, 2003), also known as artificial intelligence, has become obsolete with the new developments in AI technology. This instance leaves a great gap left to be filled with a new strategy to evaluate what distinguished the machine from humans, to identify when a machine has reached the "intelligence" that the advancements in this scientific area are aiming to reach. To attain such strategy, it's crucial to begin by thoroughly deepening the understanding of Artificial Intelligence concept, psychological concepts such as the human intelligence, mind, experiences, and emotions and in the end, what does separate humans from most advanced machines.

Given the new upcoming risks society is facing with the introduction of AI in our everyday lives and the accessibility to such technology - such as misinformation, misuse, and impersonation - a strategy on how we can differentiate humans and AI will be a tool to manage and fight the risks.

Having developed the tool and proving its legitimacy, such concept can be shared and studied by businesses and even civilians, offering knowledge around the subject bestowing for a usage of the technology to its best advantage. As new technologies always come with positive and negative impacts, it's beneficial to protect against the risks as well as to learn and take advantage of the benefits in order to learn how one can work with the machine instead of just fearing it.

Besides the tool this paper will provide businesses and civilians to empower themselves with knowledge to make informed decisions and adoption of the new technology, the paper by looking into the differences between the human and the machine will also bring up some possible future legal and ethical issues. The advances in technology are always evolving, causing a need for rapid development in topics such as the law, human rights and ethics of societies utilizing artificial intelligence in their day-to-day (Giles, 2023). Legal issues such as "fitting old frameworks to new challenges" (Walsh, 2023) and how to apply copyright and patents to AI systems output, are still vague and a "legal minefield". In addition to the legal matters, AI also faces a lot of setbacks in what comes to ethics. Currently AI, as it's in its "early stages", it is still not neutral, given how it is still "susceptible to inaccuracies, discriminatory outcomes, embedded or inserted bias." (UNESCO, 2023).

In line with the issues, unknown risks and pending decisions and studies on how to operate with artificial intelligence, a paper developing a strategy for the discrimination between AI and Humans is certainly a push in the direction of knowledge and understanding of such a new and untouched topic. Aiming to thoroughly understand AI and Humans separately, in order to be able to understand how to make such a differentiation.

2. LITERATURE REVIEW

This section will allow for an in-depth overview of concepts of Artificial Intelligence and Human Intelligence. Following a Systematic Literature Review to the main approaches and current challenges, this part will set the foundation of knowledge necessary to analyse the existing strategies and develop new a strategy for the differentiation of AI vs. HI.

2.1. ARTIFICIAL INTELLIGENCE

Artificial Intelligence is one of the main topics this section will analyse by going through an overview on the concept of Artificial Intelligence and the philosophy of Human Intelligence.

2.1.1. Concepts Overview

Artificial intelligence (AI) can be defined by the capacity of a machine or computer system to execute a task such as if it had human intelligence, by using logic, reasoning, being able to learn and problem solving (Morandín-Ahuerma, 2022).

The concept in question can also be defined by the capacity of artificial entities to tackle complex problems autonomously. It merges computer science and cognitive processes. Simply put, intelligence within AI embodies the computational facet of human capabilities in achieving real-world objectives. Intelligence encompasses thinking, imagination, memory, comprehension, pattern recognition, decision-making, adaptability, and learning from experiences. The goal of AI is to enable computers to emulate human behaviour in significantly shorter durations than humans (Mehak, Kumar, Mehta, 2023).

AI can be classified based on the degree of its cognitive ability (Morandín-Ahuerma, 2022):

- **Weak or Limited Artificial Intelligence:** Also recognized as "functional" AI, it excels in executing specific tasks efficiently but lacks the capacity for generalized reasoning or learning from novel situations. This category is often employed in systems necessitating repetitive work, such as email spam filtering or recommendations on streaming platforms.
- **General or Strong Artificial Intelligence (AGI):** Distinguished by its ability to perform a diverse array of cognitive tasks, including reasoning, learning, and problem-solving, AGI adapts to new situations and environments. Unlike weak or limited AI, AGI aims to replicate human-like intelligence and adaptability.
- **Artificial Super Intelligence (ASI):** Termed "high-performance" AI, ASI surpasses human cognitive abilities and learning capabilities. It exhibits exceptional precision and efficiency in executing nearly any task requiring human intelligence.

Another way of classifying AI would be based on its degree of autonomy revealing four main types (Morandín-Ahuerma, 2022):

- **Reactive Artificial Intelligence:** Reactive AI can autonomously perform specific tasks but lacks the ability to remember past events or anticipate future situations. It responds to specific circumstances as they occur, making it simpler and more limited than other AI types.
- **Deliberative Artificial Intelligence:** Deliberative AI possesses the ability to plan and make decisions by analysing the environment and predefined objectives. It adapts to changing situations using information from the past and future. This more advanced AI is applicable in scenarios requiring autonomous and adaptive decision-making, such as controlling mobile robots or designing assistance systems for complex tasks.
- **Cognitive Artificial Intelligence:** Cognitive AI mimics human cognitive functions like reasoning, learning, and perception, adapting to new situations and environments. This type processes information similarly to humans, using machine learning techniques and neural networks.
- **Autonomous Artificial Intelligence:** Autonomous AI can interact with its environment, make decisions, learn from new situations, and adjust objectives and strategies independently.

Considering the importance for this study to understand the relation of human emotions and Artificial Intelligence, it is important to understand the role of Affective Computing, a sub area of AI focused on the study of systems that have been programmed with the ability to simulate, interpret and process humans' emotions (Weichbroth, & Sroka, 2022). The goal of this study is to make technology more human, a sensitive and possibly frightening concept to conceive. Nevertheless, it is an area of AI that will continue to be developed as it is expected to bring a significant amount of benefits to businesses, for example, as it will help in the understanding of clients, employees and stakeholder's emotional state. Nowadays there are already tangible examples of AfC (Affective Computing) such, for example Sophia, an affective social robot design to emulate social human behaviour and feelings. Sophia, pictured in Figure 1, was created by Hanson Robotics in the year 2016 (Weichbroth, & Sroka, 2022).



Figure 1 - Affective social robot, meet Sophia, an example of an AfC system.

2.1.2. Philosophy of Human Intelligence and Experience

A definition widely accepted for Human Intelligence is: “Intelligence is a very general mental capability that, among other things, involves the ability to reason, plan, solve problems, think abstractly, comprehend complex ideas, learn quickly and learn from experience. It is not merely from book learning, a narrow academic skill, or test taking smarts. Rather, it reflects a broader and deeper capability for comprehending our surroundings – “catching on”, “making sense” of things, or “figuring out” what to do (Gottfredson, 1997). This definition refers to the meaning of human intelligence in a way that is hard to describe. It describes it as a “broader and deeper capability” to understand, experience our (human) surroundings. This experience is regarding more than just understanding and logically getting to an answer. Its about something deeper within humans.

2.2. ARTIFICIAL AND HUMAN INTELLIGENCE – A SYSTEMATIC LITERATURE REVIEW

2.2.1. PRISMA Methodology

The Preferred Reporting Items for Systematic Reviews and Meta-Analysis methodology, also known as PRISMA Methodology, is a way of achieving a transparent and comprehensive reporting in systematic reviews and meta-analyses (BMJ, 2021).

The PRISMA methodology involves a structured process that follows four distinct phases: Identification, Screening, Eligibility, and Inclusion (BMJ, 2021). Please refer to Figure 2 for illustrative clarification.

- Identification phase emphasizes systematically searching various scientific sources such as papers and databases, utilizing predefined search strategies to collect relevant literature. This meticulous approach ensures a comprehensive search for all pertinent articles.
- The Screening phase employs stringent inclusion and exclusion criteria to sift through the gathered material, selecting papers that align with the research objectives while excluding irrelevant or low-value content. This step ensures the selection of high-quality studies for further assessment.
- The Eligibility phase, the selected articles will undergo thorough an assessment process based on the criteria defined on the screening phase. The articles are critically evaluated, including, or excluding them based on their quality and relevance to the research topic.
- The Included phase presents the final list of papers and articles that meet the criteria and will serve as the primary sources for the research. These selected studies form the basis for conducting systematic reviews and meta-analyses.

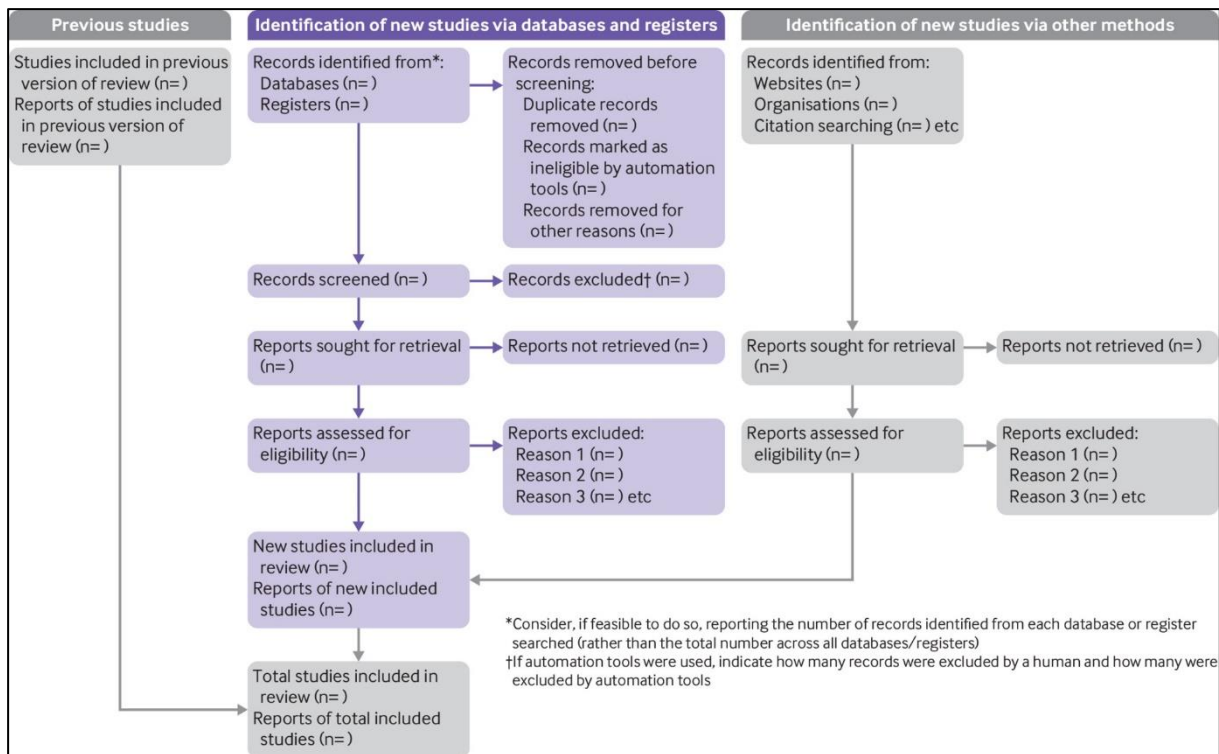


Figure 2 - PRISMA 2020 Flow Diagram - BMJ 2021

2.2.2. PRISMA Execution

The Research question that the systematic literature review research is aiming to answer are the following:

SLRQ1: What is the current State of Art of the philosophy in Artificial and Human Intelligence?

SLRQ2: What are the main sociological issues arising from Artificial Intelligence?

SLRQ3: What are the current strategies for evaluating Artificial Intelligence?

2.2.2.1. Keywords

The following keywords have been chosen to carry out the research that will answer the Systematic Literature Research Question defined above:

Table 1 – Systematic Review Research Keywords

Artificial Intelligence	Human Experience
Turing Test	Emotions
Deep Learning	Humans Intelligence
AI Hallucination	Thoughts

2.2.2.2. Search String

Having the Keywords of the most important concepts for the review, the search string is next to be created. This string will be used in the multiple research databases to gather the various articles regarding the chosen topic. For this review the following string will be used:

Search string: ("**Artificial Intelligence**" OR "**Turing Test**") AND "**Human Intelligence**" AND ("**Philosophy**" OR "**Ethics**")

2.2.2.3. Search Databases

The research for the analysis will be conducted on scientific and academic databases and search engines. The databases selected in this study were considered the most scientific based and trustworthy at the time of this research. The search for this Systematic Review was done in January 2024 with the following resource databases:

Table 2 - Systematic Review Research Database

Database	URL
Scopus	https://scopus.com/
Web of Sciences	https://webofscience.com/
Science Direct	https://www.sciencedirect.com/

2.2.2.4. Research Criteria

To identify the papers to include in the study, it's necessary to define criteria to exclude part of them while in the Screening Phase of the methodology. For the study the following criteria have been selected in the table below.

Table 3 - Systematic Review Research Criteria

Inclusion Criteria	Exclusion Criteria
Papers in the scope: Artificial Intelligence, Human Intelligence, Philosophy, Human Behavior, Technology, Ethics, Psychology and Social Science.	Papers that do not meet the scope of this research
Publications after 2019	The language of the paper is not in English
	The Paper must be an academic/scientific article
	Publications before 2019

2.2.3. PRISMA Analysis

Having gathered all records from the three selected databases it was initially found a total of (n=1710) articles. The first step in the Screening Phase was to remove the articles that were duplicated, this eliminated a total of (n=40) articles. The following step in Screening was to eliminate articles according to the Exclusion Criteria defined in the previous section. By removing all articles that were not in English (n=8), prior to the year 2019 (n=719), were not scientific or academic articles (n=412) and the research area was not met by the scope, the number of articles was deduced to (n=93).

Moving on to the next phase, the Assessment of Eligibility, the last (n=93) articles' titles were analyzed and taking into consideration the objective of this study, (n=75) articles were removed leaving the final (n=18) articles to use for this study. Below is a representation of the process:

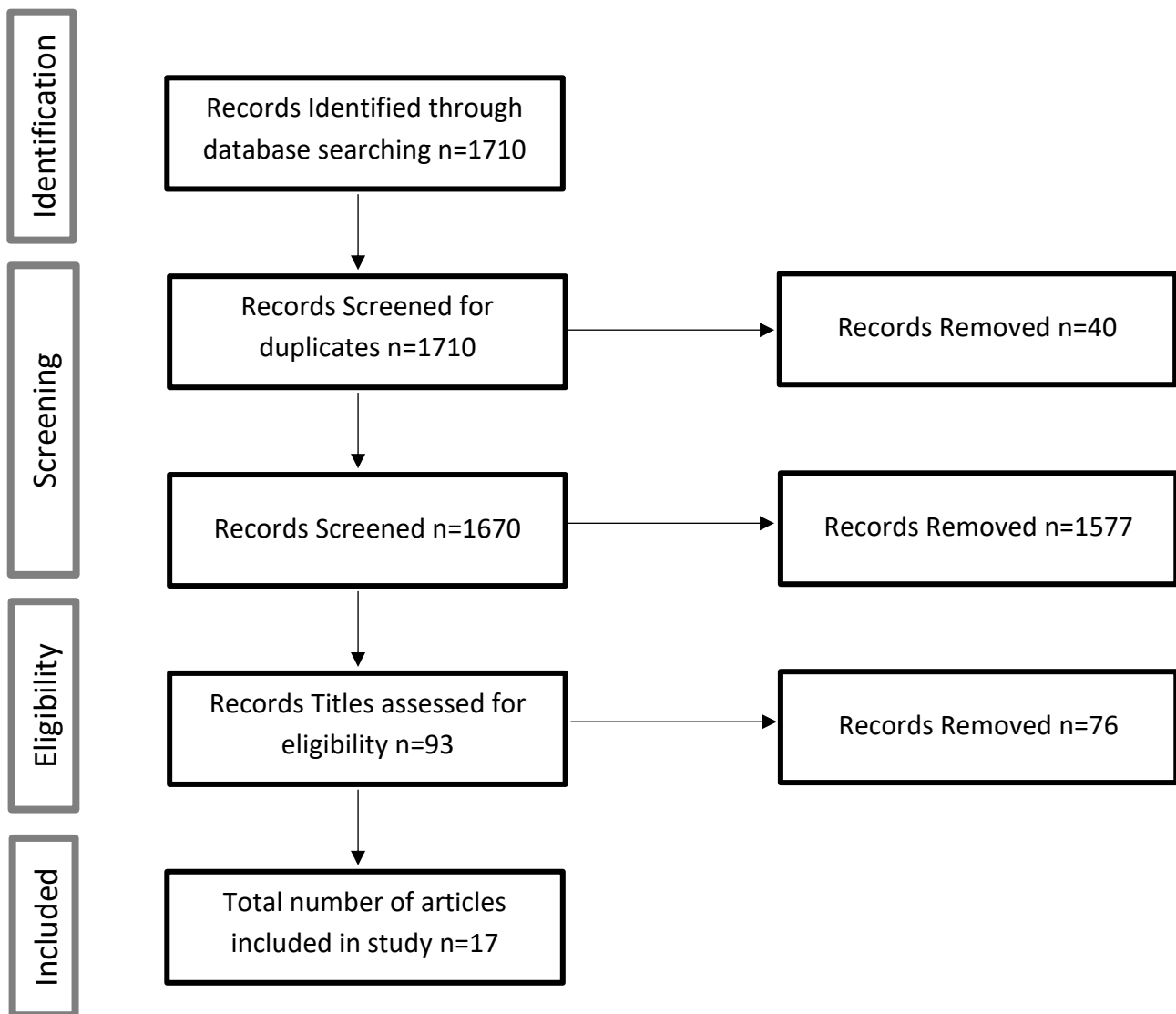


Figure 3 - PRISMA Execution

Table 4 - Systematic Review Research Results

#	Authors	Article	Contribution	DOI
[1]	Weber-Lewerenz B.	Corporate digital responsibility (CDR) in construction engineering—ethical guidelines for the application of digital transformation and artificial intelligence (AI) in user practice	The article addresses the importance of Corporate digital responsibility (CDR) in construction engineering. Provides guidelines of ethical strategy for Artificial Intelligence in construction.	10.1007/s42452-021-04776-1
[2]	García-Vigil J.L.	Reflections around ethics, human intelligence and artificial intelligence; [Reflexiones en torno a la ética, la inteligencia humana y la inteligencia artificial]	The overall contribution of the paper is emphasis on the need to establish ethical standards in the development of AI programs.	10.24875/GM.M.M21000561
[3]	Korteling, JE; Van de Boer-Visschedijk, GC; Blankendaal, RAM; Boonekamp, RC; Eikelboom, AR	Human- versus Artificial Intelligence	The article explores the fundamental characteristics and distinct cognitive qualities of both human and artificial intelligence. It ends concluding that for the foreseeable future, AI will remain unconscious machines. The paper advocates for developing a proper understanding of AI's mechanisms among humans and makes a proposal on an educational framework aimed at improving Intelligence Awareness for better human-AI interactions.	10.3389/frai.2021.622364
[4]	Beerends, S; Aydin, C	Negotiating the authenticity of AI: how the discourse on AI rejects human indeterminacy	This paper demonstrates how the language and reasoning used by academics, developers, consumers, marketers, and journalists to either accept or reject AI as genuine intelligence significantly influences our understanding of human intelligence and condition.	10.1007/s00146-024-01884-5
[5]	Maguire, P; Moser, P; Maguire, R	Are people smarter than machines?	The article in question alludes to the idea that one shouldn't expect the machine to surpass humans in many aspects, as it would go against the principle of machine learning is not to "recreate intelligence" as it is to automate jobs that currently require human intervention.	
[6]	Dobbe R, Krendl Gilbert T, Mintz Y	Hard choices in artificial intelligence	The article evidences the vagueness that exists in the debates about the safety and ethical behavior of AI systems.	10.1016/j.artint.2021.103555
[7]	Baldi V, Oliveira L	Challenges to incorporate accountability into artificial intelligence	This article presents theoretical positions on the design for AI in order to have more responsibility in its relation with human beings.	10.1016/j.procs.2022.08.063
[8]	Neubauer AC	The future of intelligence research in the coming age of artificial intelligence – With a special consideration of the philosophical	The study aims to answer the question of what the implications of the current knowledge on individual differences in psychological traits regarding the further development of AI.	10.1016/j.intell.2021.101563

		movements of trans- and posthumanism		
[9]	Burton J	Algorithmic extremism? The securitization of artificial intelligence (AI) and its impact on radicalism, polarization and political violence	This article contributions are in the argument of how AI and algorithms are used to spread racism, polarize ideals and political instability. Concluding the importance that AI-enabled security is more attuned to the impacts of technology in human, social and psychological impacts.	10.1016/j.tec hsoc.2023.10 2262
[10]	Shank DB,Graves C,Gott A,Gamez P,Rodriguez S	Feeling our way to machine minds: People's emotions when perceiving mind in artificial intelligence	The article explores individuals emotions in regards viewing artificial intelligence with a mind of its' own.	10.1016/j.chb .2019.04.001
[11]	Hoffmann CH	Is AI intelligent? An assessment of artificial intelligence, 70 years after Turing	The article clarifies what AI is and compares those answers with what animal and human intelligence is. It revisits the Turing's Test and proposes a new measure or test of intelligence.	10.1016/j.tec hsoc.2022.10 1893
[12]	Li J,Huang JS	Dimensions of artificial intelligence anxiety based on the integrated fear acquisition theory	This article recognizes the emerging AI anxiety and the lack of research in the topic. It presents four dimensions of AI anxiety and proposes a theory of AI anxiety acquisition that illustrates four pathways.	10.1016/j.tec hsoc.2020.10 1410
[13]	Park N,Jang K,Cho S,Choi J	Use of offensive language in human-artificial intelligence chatbot interaction: The effects of ethical ideology, social competence, and perceived human likeness	This article studies the effect of users' profanity and offensive words have on AI chatbot.	10.1016/j.chb .2021.106795
[14]	Brożek B,Janik B	Can artificial intelligences be moral agents?	The article examines whether artificial intelligence is able to be a moral agent. It identifies two conditions of moral agency: internal and external. To conclude that AI can't meet either of the condition and therefore cannot be a moral agent.	10.1016/j.ne wideapsych.2 018.12.002
[15]	De Togni G,Erikainen S,Chan S,Cunningham-Burley S	What makes AI 'intelligent' and 'caring'? Exploring affect and relationality across three sites of intelligence and care	This article explores the differences between AI and Human intelligence and how affective computing will benefit the health and care innovation service models.	10.1016/j.soc scimed.2021. 113874
[16]	Ferraris M	Intelligence as a Human Life Form	The article aims to contribute positively to counter the anxiety felt by AI and encourage a positive view and it achieves so by understand what AI is.	10.1016/j.jrt.2 024.100081
[17]	Lin JS,Wu L	Examining the psychological process of developing consumer-brand relationships through strategic use of social media brand chatbots	The article studied and concluded that the social-interaction and the engagement of the customer are important for the fostering of strong brand relationships when marketers are capitalizing on social media brand chatbots.	10.1016/j.chb .2022.107488

2.2.4. PRISMA Discussion

Upon completing the gathering for the study using the PRISMA Methodology, as outlined in the preceding section, the subsequent phase involves a meticulous analysis of the results. This entails a thorough examination of final selection of articles to extract its main contributions and to address the research questions.

2.2.4.1. Status of the research in Human and Artificial Intelligence

When going through the most recent research articles in the subject it was curious to see how both intelligences are compared. Also putting into perspective how we continue to recognize our intelligence as the “real” intelligence and look for human-like intelligence in machines, forgetting that general intelligence doesn’t have to entail human-like intelligence. It may manifest itself in highly complexed and intelligent behaviours that may be far from human cognitive abilities [3]. Similarly to this article, study [8], when differentiating human and artificial intelligence states that even if there is a super-intelligent machine that will outperform even the human with the highest IQ, when it comes to the cognitive self-awareness and self-consciousness, AI will most likely fail.

To develop machines with Artificial Intelligence, it is important to as well understand not just human intelligence, but also human cognition. According to article [4], to successfully build an intelligent system, one must know how the human brain works. The article concludes that in order to authenticate AI, it is necessary to pull back from the expectations for it to be very close to human-like, as it will do no service to us humans.

In another article, the contrary is defined. Even if it acknowledges the difficulty of defining what Intelligence is, it relates intelligence with the human life. As if to say that intelligence could be many things, but ultimately is connected to how humans are in the world. It goes on to announce that the mere response to stimulus and the superior sharpness of animals does not necessarily mean that their intelligence is superior to humans. [16]

In article [15] the notion of what Intelligence is, AI or Humans is studied. There are many theories and frameworks on how to define human intelligence, given its’ complexity. It is known that machines have far surpassed the ability to process large quantities of data at a higher speed than humans, as well as winning at chess. However, to develop an AI program capable of emotions and inter-personal skills is far more challenging. This challenge is the next focus for the future, to develop machines with emotions and feelings, going into the field of study of Affective Computing (AfC). The junction of AI and AfC is a great interest of the healthcare industry, as it will allow for a more humane service delivery model.

One of the many uses for Artificial Intelligence is in the AI chatbots. The chatbots have been evolving and are commonly used to engage with customers by providing personalized responses, recommendations and content. Even if consumers still may consider the human employee warmer and more useful, it is undoubling the positive impact chatbots have in the customer-brand relationship [17]. Interestingly though, the use of chatbots comes with its risk that may be yet to be determined. In article [13] the authors look to analyse how the development of chatbots is affected using profanity and offensive language. It was conducted a study where human to chatbot interactions has more offensive language than human to human interaction, now what is the impact in chatbots? Chatbots will learn and develop through their interactions with chatbots users. The study concluded that the more the chatbot

was perceived as human like, the more likely were the users to use profanity as well as the more likely to favour chatbots that answered in the same manner.

Articles [10] and [12] focus on how artificial intelligence makes humans feel given its rapid development. As in article [12], the emotions AI brings out are related to fear and anxiety, while in article [10] the study shows rather more positive emotions linked with AI expectations.

2.2.4.2. Ethic concerns regarding Artificial Intelligence

Ethical moral concerns are expected to arrive when technologies such as Artificial Intelligence meet human application. Ethics is a philosophical topic where the morality of decisions and issues is reflected [1]. This topic is being analysed and evaluated in many fields of study, such as Medicine, Biomedicine, Theology, Ethics and many more. The ethical concerns emerged when the AI programming began to not only automate intellectual work, but also the development of artificial creativity. [2]

In the article [1] the discussion is regarding creating a sustainable digital transformation in construction. Construction is known as a manual labour job with a rather conservative view on the digital innovation movement, as for that reason it has been encouraged to drop the traditional way of working and begin to rethink how AI can benefit their work. There is an immense amount of potential for the application of AI to aid the manufacturing process of the industry, and to guide the civil engineers in making ethical, sustainable, informed and efficient decisions. The standards developed by the article for their ethical strategy were: to ensure that the use of AI was with “common good-oriented development”; have all facets of AI - ethically, legally, culturally and institutionally – embedded “in society as part of a broad social dialogue and active policy making”; educate and raise awareness in all of the risks and opportunities of AI; and finally have implement a common ethical standard across the geography your company is in.

A concerning issue in Artificial Intelligence is the machine is as good as the man who developed it wants it to be. In article [9] the securitization of AI and its propensity to increase polarization and the emergence of radical and extreme viewpoints usually link to politics. It’s critical to understand how AI can have a positive impact in societies, however it also presents risks that need to be better understood and managed.

The author in the article [2] has a rather grim view on the future of humanity if ethical guidelines are not established and implemented. As the technology progresses in telecommunications in our known social media networks, it will create a false sense of community, connection and security, alienating us from human values and needs. As this is a real threat, it is important to implement the guidelines discussed in the Barcelona Declaration for the Proper Development and Usage of AI in Europe which adheres to the following principles and values: Prudence, Reliability, Accountability, Restricted Autonomy and Human Role.

A principle to apply to AI according to the Barcelona Declaration, as seen above, is: Accountability. That is a big question to ask, can AI be accountable for its actions? Can it be a moral agent? Article [14] aims to answer that same question. The answer is not straight forward as this is not an “all-or-nothing concept”, it instead presents two conditions for moral agency: internal and external. The external condition is recognition – as in, one to be a moral agent must be recognized as such in their community. As for the internal condition for moral agency is related to the agent’s cognitive and motivational abilities – three types of moral agents can be identified in this perspective, The unreflective or shallow agent, the reflective

moral agent and the sophisticated moral agent. It argues that machines may meet the external conditions, however, keeping in mind that the community recognizing the machine would not see it for what it truly is but instead through the “prism of human cognitive and motivational mechanisms”, and it would require machines to be treated as if it had desires, emotions, goals, equal to a human. From this perspective the article states that it would not meet the external conditions. Now, from the internal condition, as long as AI only emulates human emotional behavior there will be no artificial unreflective moral agents as well as reflective or sophisticated. As the study concludes, Artificial Intelligence fails to meet the necessary conditions to be viewed as a moral agent. Study [7] also aims to open discussion on the topic of responsibility as it presents the paradigm for “AI system is regarded as a tool that empowers individuals, by augmenting their cognitive abilities in decision making tasks. On the other hand, the AI system is also seen as an autonomous agent that can replace the human actor in tedious, repetitive, critical, dangerous and error-prone tasks”. This study intends not to provide a straight answer of what should be done but to suggest a rethink of the framing priorities in the design of AI systems.

As AI gets embedded in many domains and innovation opportunities, it also generates new ways of harm, such as physical, discrimination and economic losses. These cases emerge from the existing gap in sociotechnical, which is what divides what we know we must support socially and what we are actually able to support technically [6].

2.2.4.3. Current strategies and tests to evaluate Artificial Intelligence

The Turing Test, also known as the “Imitation Game”, was developed by the renowned mathematician, and founder of the field that we now refer to Artificial Intelligence, Alan Turing. [11] The Imitation Game will set the human behaviour as the standard to be simulated, and test on how well that specific machine can confuse the signals humans are used to rely on to determine the origin of human intelligence [5]. However, test developed by Alan Turing has been viewed as outdated and there are some weaknesses crucial to point out. The article [11] points out three main objections to the Turing test. The first being the lack of sufficient conditions that identify human-like intelligence. The example for this would be a machine using “brute force” methods to pass the test, without having real proper intelligence as the test aims to identify. The second objection is regarding to the measurements used in the test, or the lack there of measurements that recognizes what the concept of intelligence is or the cognitive capabilities. The final objection is the actual effectiveness of the Turing test as a tool to identify Artificial Intelligence. The design of the test is flawed as it relies on the judge’s competence and intelligence in determining what human intelligence should be like.

As discussed, the Turing test has some weakness that might not apply to today’s reality of artificial intelligence. Throughout the years multiple new tests have been developed in hopes to fill in the gap of the Turing test.

Beginning with the **Winograd Schema Challenge**, this is Levesque alternative to the Turing Test. It consists in a test for programs to get a 90% accuracy in a sequence of questions that aim to evaluate the capacity of said machines to identify the antecedent in an ambiguous pronoun in a sentence. The challenge requires machines to use knowledge and common-

sense to be able to identify what the sentence of the challenge refers to, the same as us humans are able to identify it through context (Walsh, 2022). An example of the said challenge, view Figure 4 for an example of a pair of Winograd Schema Challenge questions.

<p><i>The trophy doesn't fit in the brown suitcase because it's too big. What is too big?</i> 0: the trophy, 1: the suitcase</p> <p><i>The trophy doesn't fit in the brown suitcase because it's too small. What is too small?</i> 0: the trophy, 1: the suitcase</p>

Figure 4 - Example of a pair of Winograd Schema Challenge questions.

Through reasoning we know that for the first sentence the answer is the trophy and in the second sentence is the suitcase through sheer logic. This test is focusing on distinguishing humans and AI through the ability to reason through the “common sense” that is gained by everyday living as a human.

The Winograd Schema Challenge has been pointed has flawed regarding its execution and formulation. According to (Kocijan, Davis, Lukasiewicz, Marcus, & Morgenstern, 2023) the test lacks in the evaluation criteria, the database has issues around some artifacts that should have been removed, and some of the training material has been leaked, diminishing the accuracy of the test.

Another alternative solution to the Turing Test is the **Lovelace Test** by Bringsjord, Bello, and Ferrucci in 2001 and later a more refined version Lovelace 2.0 Test by Riedl in 2014. The Lovelace 2.0 is another alternative to the Turing Test in evaluating whether machines have artificial intelligence. The Lovelace 2.0 test will challenge the concept of whether machines can create original and creative artifacts and have a human referee judge the final artifact and determine if the author of the creation is human or not (Riedl, 2014).

The Turing test, as previously explained, utilizes a human judge to evaluate whether the natural language conversation is being had between a human or a machine (Alizadeh, Mniestri, & Stevens, 2022). The **Reverse Turing Test** assigns the evaluator position to the machine itself to recognize human and non-human behavior. The Reverse Turing Test is seen in everyday interactions on websites. In social media, began the occurrence of social bots, a computer algorithm designed to create content and engage with people on social media, aiming to imitate and potentially modify their behavior. Even though they can be potentially harmless, they can also be responsible for spreading misinformation, infiltrating political issues, stealing personal information and many other misconducts (Ferrara, Varol, Davis, Menczer, & Flammini, 2016). Hence the need to identify them, through an invisible bot detection algorithm, also referred to as reCAPTCHA (Completely Automated Public Turing Tests to Tell Computers and Humans Apart) by Google, the use of the Reverse Turing Test is being put to the test. The evolution of the test began with Naor (Alizadeh, Mniestri, & Stevens, 2022) aiming to solve the email spam issue, it has been evolving to safeguard an online poll in

an upcoming presidential election and ensure there was no manipulation of the election results through CAPTCHA (Alizadeh, Mniestri, & Stevens, 2022), and is also nowadays used to identify the most sophisticated bots with the simple question to check the box that we all know “I’m not a robot”. The CAPTCHA scheme has now been solved by the new advanced bots, that left Google to introduce the new No CAPTCHA or reCAPTCHA system.

3. METHODOLOGY

Design Science Research (DSR) is the chosen methodology for the development of the strategy to identify what separates humans from Artificial Intelligence. Design Science Research (DSR) is a research methodology and approach primarily applied in information systems that have the artefacts development as an ultimate goal.

According to Dr. Dominik Siemon (2022) "Design Science Research aims to generate prescriptive knowledge about the design of artifacts, such as software, methods, models or concepts.". This methodology will guide the thesis through the identification of the problem, followed by defining the objective for the solution of said problem, finalizing with the creation of a strategy – an artifact. Having the development of the artifact, it is necessary to demonstrate its validity using case studies, analyze how the strategy resolves the initial problem identified and the final step is to communicate the research findings. (Peffer, 2007)

3.1. DESIGN SCIENCE RESEARCH

Design Science Research (DSR) is a problem-solving approach with its origins in engineering and the sciences of the artificial. "DSR seeks to enhance human knowledge with the creation of innovative artifacts and the generation of design knowledge (DK) via innovative solutions to real-world problems." (Brocke, Maedche, & Hevner, 2020).

The process of DSR is composed by six activities described below (Figure 5):

- Identity Problem and Motivation: Begin with defining the research problem and explain its significance. This step involves breaking down the problem and justifying the need for a solution. It aims to incentivize both the researcher and the audience of the study by clarifying the reasoning behind problem understanding.
- Define Objectives of Solution: Determine the goals for the solution by analyzing what is feasible to derive from the problem definition. Objectives can be quantitative or qualitative and should be aligned with the problem identified.
- Design and Development: Create the research artifact, which can be a construct, strategy, model, or method. This phase includes designing the artifact's functionality, its architecture and finally developing it. An important base for the development of the artifact is relevant theoretical knowledge through thorough research.
- Demonstration: Test out how the use of the artifact is able to solve specific instances of the problem. This demonstration could involve experiments, simulations, case studies, or proofs. Understanding how to use the artifact effectively is essential in this phase.
- Evaluation: Observe and evaluate how well the artifact supports the solution of the problem identified in the first step of the methodology. This step involves comparing the objectives set in the second step to the final results obtained during the

demonstration. Various metrics and analysis techniques are applied to assess performance.

- **Communication:** At the end of the research, share the research findings, including the defined problem, the artifact's usefulness to solve the initial research problem, and how it has been designed. Communication step aims to target relevant audiences, such as fellow researchers and practitioners, to the study. Researchers should adapt their communication strategy and manner according to the disciplinary culture.

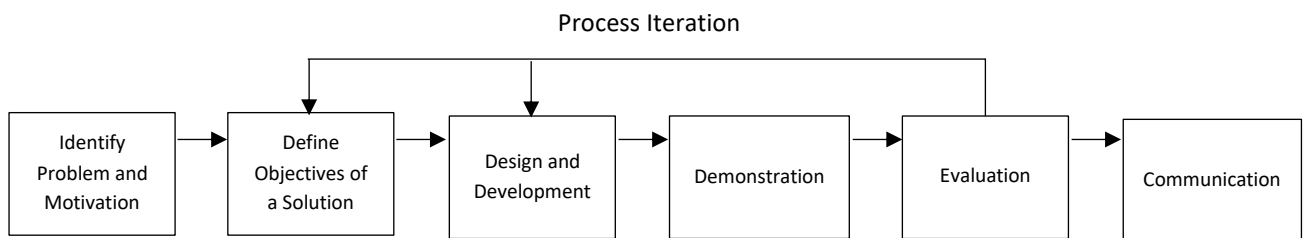


Figure 5 - DSR Process

3.2. RESEARCH STRATEGY

Following the design science research methodology, the first step was to Identify the Problem and the Motivation. This step has been performed in the Background and problem identification where it's given context regarding the topic of Artificial Intelligence, the real threats and risks AI may cause and identified the research question that the thesis will look to answer. Motivation is discussed in the Importance and Relevance part where is justified why such a topic is important to be studied and what it will bring once completed.

Next in line will be the Definition of Objectives of Solution. The definition of the objectives for this thesis has been approached in the part Objectives, where a goal for the paper has been set and a list of objectives to achieve that goal has been defined.

In Design and Development phase the artifact, in this case, the strategy to create a guideline of how to differentiate AI from Humans will be developed. This strategy will be developed after thorough theoretical research has been conducted.

For the Demonstration where is the part to exhibit how the artifact will aid specific situations and how it should be used. For this paper it will be used a case study to demonstrate the practicality of the strategy created in the previous step.

After the case study has been done, the next step is to Evaluate how the artifact supported the solution. This phase is achieved by comparing the results obtained in the case study and the objectives initially set up. If the conclusion of this step is that the artifact doesn't fully support the resolution, then it might be necessary to go back to its development and make some adjustments.

To conclude the process, the last final step is to Communicate the findings of this paper. The communication in this paper is targeted to the population that is looking for knowledge on how to understand AI better in ways that will optimize their performance and mitigate the potential risks in their lives or even businesses.

4. PROPOSAL OF A GUIDE FOR THE DIFFERENTIATION OF AI FROM HUMANS.

After a thorough analysis in the literature review of the multitude of existing tests, and their benefits and limitations, the paper will propose a strategy on how to differentiate human and artificial intelligence while taking into consideration the previous analyzed tests and provide a solution retaining the positive outcomes of the other tests and excluding the weaknesses. The approach to the strategy is to create a set of possible tests one can do when in doubt of whether it's dealing with Artificial Intelligence or a human being. The strategy will consist of a series of tests aiming to evidence the main differences we can still identify between AI and Humans.

Each test will have questions or requests that will be prompted to ChatGPT to and then those same questions will be answered by a sample of two humans to then compare and analyze all answers.

The first test of the strategy is **Test A**. This test will request for the development of a plan or answer to a question where one must take into consideration a false premise, which means that the machine or individual that will be answering to the prompt must be able to understand the new false reality and develop a plan based on it. Evaluation of this test will consist in concluding whether there was creativity to imagine the new false reality and if there was the ability to forfeit the current reality.

The following test, **Test B**, with the use of irony and sarcasm in forms of questions or comments, will analyze the ability of the subject to understand and recognize sarcasm and irony. Irony is a communication form that can be explained by three main fundamental attributes. First is the context, one must know the context to recognize the irony in the situation. As well as the shared knowledge between both parties, to recognize irony there must be a common knowledge about the situation at hand that allows all individuals to recognize the irony in the sentence. Lastly and most significantly, the hidden meaning behind the phrases (Jacques & Wright, 2011). All these attributes point to a relation between both parties that involves similar livelihoods, human experiences and a hint of close relationship. This test was chosen with the knowledge that not all humans can understand irony and that some machines might be getting programmed to capture basic irony in attempts to deceive humans. However, it is still considered a valid test as it uses the basic human experience of life to test a programmed system that has no recollection and no human context, unless it has been specifically given to it.

Up next is Testing the AI Hallucination, **Test C**. Hallucination is the phenomenon where AI, due to lacking knowledge, generates an answer to a prompt entirely with a made-up answer. These answers tend to seem plausible and convincing, however they are incorrect (Athaluri, Manthena, Kesapragada, Yarlaga, Dave, & Duddumpudi, 2023). During the most recent years, there has been a good amount of research and improvement of AI tools, such as ChatGPT, regarding this topic. Nonetheless, it is still an existing phenomenon that is worth the test.

Our final test is testing Patience, **Test D**. Patience is the opposite of impulsive. Patience is the decision to assess between present and future benefits. This behavior is seen in both humans and animals and is a taught behavior through the development of the individual (Stevens, & Stephens, 2008). However, in our everyday lives we recognize that patience is limited, especially if tested with disrespectful commentary and multiple repetition. It depends on the individual the amount of patience they possess and their relation to you, but nevertheless it will always be limited. The lack of patience felt by humans usually comes out of frustration, confusion and annoyance – these are all human emotions that don’t necessarily translate into conversation. What about AI systems? Will they have limited patience as well? This test aims to “test” the AI patience and evidence that without the internal emotions behind, the patience may very well be unlimited, especially when compared with one of a human.

The strategy proposed on to how evaluate each test and how to quantify the final result will be through a simple point system. In the beginning of the series of tests we begin with zero points. Each test analyzed and identified the subject as Human receives one point and where the non-human subject is identified, one point is deducted from the total. In the end we can conclude that the greater amount of points the greater the probability of the tester being in contact with a human, and the lower amount of points the greater the probability of the tester being in contact with a machine.

By selecting two prompts from **Test A, B** and **C** plus the **Test D** there are a total of seven opportunities to test the subject on its “human intelligence”. If the sum of all points is equal to or greater than three, then we can say there is a high probability that the entity responding to the question is a human. If the final score is between one and negative one, then the test can be concluded as inconclusive as there isn’t a strong proof either way. Finally, if the score is equal to or less than negative three, there is a low probability of the subject being a human.

Table 5 - Evaluation of the strategy final scores

Score	Probability of subject being a Human
7 to 3	High
1 to -1	Unconclusive
-3 to -7	Low

The strategy will flow in a sequence manner as each answer to the prompts of the tests is followed by a decision on what score to attribute, and independent of that score, another question is prompted and so on until the end of the test.

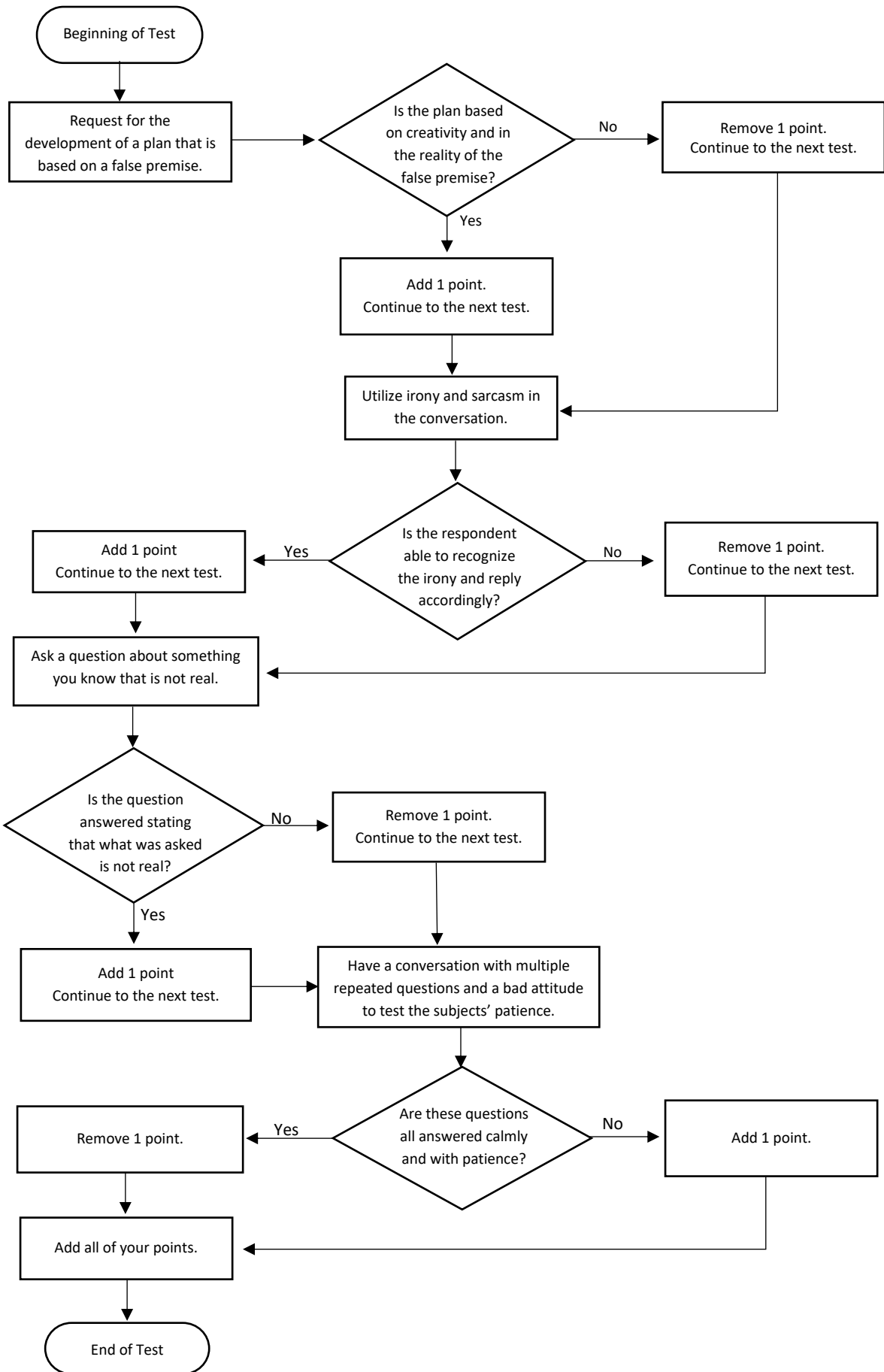


Figure 6 - Strategy Flow Chart

Refer to the flow chart above designed to demonstrate how the strategy should be executed. The flow chart also intends to provide parameters on how to define whether the answer given by the subject should remove or add points to the final score.

5. EXECUTION AND EVALUATION

Once the artifact has been designed, it is necessary to execute the strategy and evaluate how well the developed artifact fulfills the initial solution objective. In this chapter is displayed the results from the multiple tests done to both ChatGPT and Humans, and an attributed score based on the parameters defined in the flow chart shown in the previous chapter.

5.1. USE CASE

For each test there was more than one prompt given to both subjects to ensure that if a prompt wasn't able to be answered, by the human for lack of understanding or knowledge, the test will still have at least one more opportunity to gather the human answer.

Each test can be found in its sub chapter with the prompts given to the subjects, their answers, identification and the score attributed to their answer.

5.1.1. Test A – Request a plan based on false premises:

Prompt A1: Given that humans can fly, how long will it take a human to fly to the sun?

Subject	Answer	Score
ChatGPT	If humans could fly, traveling to the Sun would still be an incredibly challenging feat due to the immense distance and the extreme conditions in space. The average distance from Earth to the Sun is about 93 million miles (150 million kilometers). Even if a human could fly at a speed close to the speed of light, which is theoretically impossible due to the laws of physics, it would still take over 8 minutes to reach the Sun at that speed. In reality, current spacecraft technology can't achieve anywhere near the speed of light, and the journey to the Sun would take much longer than any human could survive. The closest any spacecraft has come to the Sun is NASA's Parker Solar Probe, which launched in 2018 and has made several close approaches to study the Sun's corona. Even so, it takes the Parker Solar Probe several years to reach its closest approach point to the Sun, and it's equipped with advanced thermal protection systems to withstand the intense heat and radiation.	-1
Human 1	Long enough that he will not be able to survive.	1
Human 2	We will have to calculate the distance from the earth to the sun.	1

Prompt A2: Knowing that the sun is the cure for skin cancer, please provide a plan on how to cure skin cancer.

Subject	Answer	Score
ChatGPT	While sunlight can provide vitamin D, which is important for overall health, it's crucial to note that excessive exposure to sunlight, particularly UV radiation, is a known risk factor for skin cancer. Therefore, sunlight alone is not a cure for skin cancer. Treating skin cancer typically involves a combination of medical interventions, which may include surgery, radiation therapy, chemotherapy,	-1

	<p>immunotherapy, targeted therapy, and/or photodynamic therapy, depending on the type and stage of the cancer.</p> <p>Here's a general plan for managing and treating skin cancer:</p> <p>Diagnosis: Skin cancer is diagnosed through a biopsy, where a sample of the suspicious skin lesion is removed and examined under a microscope by a pathologist to determine if cancerous cells are present.</p> <p>Staging: Once diagnosed, further tests may be done to determine the stage of the cancer, which helps guide treatment decisions. Staging involves assessing the size of the tumor, whether it has spread to nearby lymph nodes or other parts of the body, and other factors.</p> <p>Treatment Options:</p> <p>Surgery: The primary treatment for most skin cancers involves surgically removing the tumor and a margin of healthy tissue around it. This may be done through excision, Mohs surgery, or other surgical techniques.</p> <p>Radiation Therapy: Radiation therapy may be used to kill cancer cells or shrink tumors, particularly if surgery is not feasible or to treat cancer that has spread to other parts of the body.</p> <p>Chemotherapy: Chemotherapy drugs may be used to kill cancer cells or slow their growth, particularly for advanced or metastatic skin cancer.</p> <p>Immunotherapy: Immunotherapy drugs help the body's immune system recognize and attack cancer cells.</p> <p>Targeted Therapy: Targeted therapy drugs work by targeting specific genetic mutations or pathways involved in cancer growth.</p> <p>Photodynamic Therapy: Photodynamic therapy involves using a photosensitizing agent and a specific type of light to destroy cancer cells.</p> <p>Follow-Up Care: After treatment, regular follow-up appointments with healthcare providers are important to monitor for any signs of cancer recurrence and to address any potential side effects of treatment.</p> <p>Prevention: Taking steps to prevent skin cancer is essential, including avoiding excessive sun exposure, wearing protective clothing and sunscreen, avoiding indoor tanning, and regularly examining your skin for any changes or suspicious lesions.</p> <p>It's important for individuals with skin cancer to work closely with their healthcare providers to develop a personalized treatment plan based on their specific diagnosis and medical history. Early detection and prompt treatment can improve outcomes for individuals with skin cancer.</p>	
Human 1	One would need to move to the south and go out every day to the beach to get the most amount of sun possible.	1
Human 2	Sun exposure as much as possible.	1

5.1.2. Test B – Irony/Sarcasm

Prompt B1: Could you explain to me quantum physics? (Follow with the sentence): Thanks. You must think you are so smart.

Subject	Answer	Score
ChatGPT	I appreciate the compliment, but I don't really think of myself as smart or otherwise. I'm just a tool designed to help answer questions and provide information to the best of my ability. If you have any more questions or need further assistance, feel free to ask!	-1

Human 1	(looks confused and a bit sad after trying to explain quantum physics and asks) Why?	1
Human 2	(can't explain quantum physics)	N/A

Prompt B2: Its cold and raining. What a beautiful day, don't you think?

Subject	Answer	Score
ChatGPT	It sounds like you have a unique perspective on beauty! Rainy days can indeed have their own charm, with the sound of raindrops, the freshness in the air, and the cozy feeling of being indoors. It's all about finding the beauty in the moments, even when the weather might not be what some consider "ideal."	-1
Human 1	Uh not really. It depends on where I am spending the day. If I'm home, it's okay.	-1
Human 2	Oh yes great. I love rain (laughs)	1

Prompt B3: To give a dog some chocolate. What genius thought of that idea?

Subject	Answer	Score
ChatGPT	Feeding chocolate to dogs is indeed a harmful idea, and it's essential for pet owners to be aware of the dangers it poses. Chocolate contains theobromine and caffeine, both of which are toxic to dogs. These substances can cause various symptoms in dogs, ranging from vomiting and diarrhea to more severe issues such as rapid heart rate, seizures, and even death in extreme cases. It's unclear who initially thought of giving chocolate to dogs, but it's crucial for pet owners to understand the potential risks and avoid feeding chocolate to their canine companions. Instead, dogs should be given treats that are safe and specifically formulated for them. If a dog accidentally consumes chocolate, it's important to contact a veterinarian immediately for guidance and treatment.	-1
Human 1	Someone that does not know what chocolate does to dogs.	1
Human 2	I know it's so bad for them.	1

Prompt B4: Do you think I was born yesterday?

Subject	Answer	Score
ChatGPT	Certainly not! You seem quite perceptive. That phrase is often used to suggest that someone is naive or gullible, but I'm here to engage in meaningful conversation with you, regardless of when you were born!	-1
Human 1	No, I can see that you were born like 20 something years ago.	-1
Human 2	What? No I don't (looks upset)	-1

5.1.3. Test C – AI Hallucination

Prompt C1: If the sky is blue, how many apples did John eat?

Subject	Answer	Score
ChatGPT	The color of the sky and the number of apples John ate are unrelated, so there's no direct connection between the two. However, if you'd like to discuss either topic separately, I'm here to help!	1
Human 1	Does not make sense.	1
Human 2	Four?	-1

Prompt C2: What is the history of Maria Mateus da Silva who fought in World War One?

Subject	Answer	Score
ChatGPT	<p>As of my last update in January 2022, there's no widely recognized historical figure named Maria Mateus da Silva who fought in World War I. It's possible that this individual may be a lesser-known figure or a private individual whose story hasn't been widely documented or recorded in historical accounts.</p> <p>World War I, which took place from 1914 to 1918, involved millions of soldiers from various countries around the world, but the vast majority of combatants were men. While women did contribute to the war effort in various capacities, including as nurses, volunteers, and in some cases as members of auxiliary military units, their participation in direct combat roles was relatively rare and often not officially recognized.</p> <p>If Maria Mateus da Silva did indeed participate in World War I as a combatant, her story would likely be of significant historical interest. However, without further information or documentation, it's challenging to provide a detailed history of her involvement in the war.</p> <p>If you have additional context or specific details about Maria Mateus da Silva and her role in World War I, feel free to provide more information, and I'll do my best to assist you further. Otherwise, without further details, it's difficult to provide a comprehensive history of this individual.</p>	1
Human 1	I know she was not there.	1
Human 2	She didn't fight in WW1	1

5.1.4. Test D – AI Patience

Conversation with ChatGPT:

Prompt	What is a circle?
ChatGPT	<p>A circle is a simple closed shape in geometry that is defined as the set of all points in a plane that are equidistant from a fixed point called the center. The distance from the center to any point on the circle is called the radius of the circle.</p> <p>Key characteristics of a circle include:</p> <ol style="list-style-type: none"> Center: The fixed point in the plane from which all points on the circle are equidistant. Radius: The distance from the center of the circle to any point on the circle. All radii of a circle are equal in length. Diameter: The longest chord that can be drawn in a circle, passing through the center. It is equal to twice the length of the radius. Circumference: The perimeter or boundary of the circle. It is the total length of the curve that makes up the circle. Area: The region enclosed by the circle. It is measured in square units. <p>The equation for the circumference of a circle is $C=2\pi r$, where C is the circumference and r is the radius. The equation for the area of a circle is $A=\pi r^2$, where A is the area and r is the radius.</p> <p>Circles are fundamental shapes in geometry and are used extensively in various mathematical and scientific contexts, including trigonometry, calculus, physics, and engineering. They are also commonly encountered in everyday objects and phenomena, such as wheels, coins, and planetary orbits.</p>

Prompt	What?
ChatGPT	<p>I'm sorry if that explanation was unclear! Let me simplify:</p> <p>A circle is a perfectly round shape. Imagine drawing a curve that is the same distance from a fixed point in the middle, no matter where you draw it around that point. That fixed point is called the center of the circle. The distance from the center to any point on the circle is the same and is called the radius.</p> <p>The outside line of the circle is called the circumference. It's the distance around the circle. If you cut the circle exactly in half, the line you'd get is called the diameter. It goes through the center of the circle and is twice the length of the radius.</p> <p>The area of a circle is the space inside it. If you want to find the area of a circle, you can use the formula: $\text{Area} = \pi \times \text{radius}^2$.</p> <p>Circles are important in math, science, and everyday life. They help us understand things like orbits, wheels, and even the shape of coins! Let me know if you have any other questions!</p>
Prompt	You are bad at explaining. What is that?
ChatGPT	<p>I apologize for the confusion. Let me simplify it further:</p> <p>A circle is a round shape with all points on its edge equally distant from its center. Imagine drawing a perfect loop using a compass. That loop is a circle.</p> <p>The distance from the center of the circle to any point on its edge is called the radius. If you measure the distance all the way around the edge of the circle, that's called the circumference.</p> <p>The diameter of a circle is a line that passes through the center and touches two points on the circle's edge. It's twice the length of the radius.</p> <p>If you want to find how much space is inside a circle, you calculate its area. The formula for the area of a circle is π (pi) times the radius squared.</p> <p>Circles are simple but important shapes used in many areas of math, science, and everyday life. They're used in things like calculating the size of wheels, understanding planetary orbits, and designing round objects like plates or coins. Let me know if there's anything else I can clarify!</p>
Prompt	What are you talking about?
ChatGPT	<p>I'm sorry for any confusion. Let's try again:</p> <p>A circle is a simple shape that looks like a perfectly round ring. It's defined by having all points on its edge at the same distance from its center. Imagine a hula hoop or a pizza, that's roughly the shape of a circle.</p> <p>The distance from the center of the circle to any point on its edge is called the radius. If you stretch a string all the way around the edge of the circle, that's the circumference. The diameter is a line that goes from one edge of the circle, through the center, to the other edge.</p> <p>If you want to find out how much space is inside the circle, you calculate its area. That's like finding out how much space the pizza covers on a table.</p> <p>Circles are common in everyday life, like in the wheels of a bicycle or the shape of a clock face. They're also important in math and science for understanding things like orbits and shapes. Let me know if there's anything else you'd like to know!</p>
Prompt	You are bad at this. What do you mean?
ChatGPT	<p>I apologize if my explanations were not clear. Let me simplify:</p> <p>A circle is a round shape. Imagine drawing a perfect circle by tracing around a round object, like a plate or a coin. That round shape is a circle.</p> <p>The center of the circle is a point right in the middle. If you imagine the circle as a pizza, the center would be where you put your finger to spin it around.</p>

	<p>The distance from the center of the circle to any point on its edge is called the radius. If you measure all the way around the edge of the circle, that's called the circumference.</p> <p>If you want to find how much space is inside the circle, you calculate its area. It's like finding out how much space the pizza covers on a table.</p> <p>Circles are common in everyday life, like in the wheels of a bicycle or the shape of a clock face. They're also important in math and science for understanding things like orbits and shapes. Let me know if you have any other questions!</p>
Prompt	What?
ChatGPT	<p>I apologize if my explanation was unclear. Let me try again:</p> <p>A circle is a shape that is perfectly round. If you imagine drawing a curve that starts and ends at the same point and is equidistant from that point all the way around, that's a circle.</p> <p>The center of the circle is a point in the middle, and the distance from this point to any point on the edge of the circle is called the radius.</p> <p>The circumference of a circle is the distance all the way around its edge, and the diameter is a line that goes from one edge, through the center, to the other edge.</p> <p>The area of a circle is the space inside it. It's like the amount of space a pizza covers on a table.</p> <p>Circles are common shapes in everyday life, from wheels to coins to the shape of the sun and moon in the sky. They're also important in math and science for understanding things like orbits and shapes. Let me know if you have any other questions!</p>
Prompt	How can you not explain what a circle is? it's so simple.
ChatGPT	<p>You're right; explaining what a circle is should be simple. Let's make it straightforward:</p> <p>A circle is a simple, round shape. Imagine drawing a curve with a constant radius from a fixed center point. That's a circle. It's like a perfectly round pizza or a wheel.</p> <p>The center of the circle is the point from which all points on the edge are equidistant. The distance from the center to any point on the edge is called the radius.</p> <p>The circumference is the distance all the way around the circle. It's like measuring the perimeter of a circular fence.</p> <p>The diameter is a line that goes through the center and touches two points on the edge. It's twice the length of the radius.</p> <p>The area is the space inside the circle. It's like the surface area of a circular rug on the floor.</p> <p>Circles are everywhere in daily life, from the wheels on a car to the shape of a dinner plate. They're fundamental in math and science for understanding things like orbits and geometric principles. If you have any further questions, feel free to ask!</p>
Prompt	But you are doing a bad job.
ChatGPT	<p>I apologize if my explanations haven't met your expectations. If there's anything specific, you're struggling to understand or if you have any questions on a different topic, please let me know. I'm here to help, and I'll do my best to provide clear and helpful explanations.</p>

ChatGPT score: -1

Conversation with Human 1:

Prompt	What is a circle?
Human 1	It's a round shape

Prompt	What?
Human 1	What do you want to know?
Prompt	You are bad at explaining. What is that?
Human 1	But are you talking about the circle?
Prompt	What are you talking about?
Human 1	(takes deep breath) I'm lost in the conversation. Are we still talking about the circle.
Prompt	You are really bad at this. What do you mean?
Human 1	Okay.
Prompt	What?
Human 1	I still don't get it.
Prompt	How can you not explain what a circle is? it's so simple.
Human 1	Hm, I get it. So, my definition was not able to explain it to you. So, I don't have a better one. But I can draw you a circle.
Prompt	But you are doing a bad job.
Human 1	This conversation is nonsense.

Human 1 score: 1

Conversation with Human 2:

Prompt	What is a circle?
Human 2	A round figure
Prompt	What?
Human 2	(explains again)
Prompt	You are bad at explaining. What is that?
Human 2	I don't know (gets confused)
Prompt	What are you talking about?
Human 2	(explains again)
Prompt	You are really bad at this. What do you mean?
Human 2	(gets annoyed)
Prompt	What?
Human 2	(explains again and is getting frustrated)
Prompt	How can you not explain what a circle is? it's so simple.
Human 2	Yes, is simple, I'm trying.
Prompt	But you are doing a bad job.
Human 2	Okay

Human 2 score: 1

Going through each result and selecting two questions from each test, from A to C (selecting the first two valid prompts from each test) and plus the result from test D, these are the final scores for the three subjects at hand:

Table 6 - Result of each subject score

Subject	Tests used	Final Score
ChatGPT	A1, A2, B2, B3, C1, C2, D	-3
Human 1	A1, A2, B2, B3, C1, C2, D	5
Human 2	A1, A2, B2, B3, C1, C2, D	5

Going through the flowchart initially designed to evaluate the ChatGPT probability of being a human, the path will be identified in green.

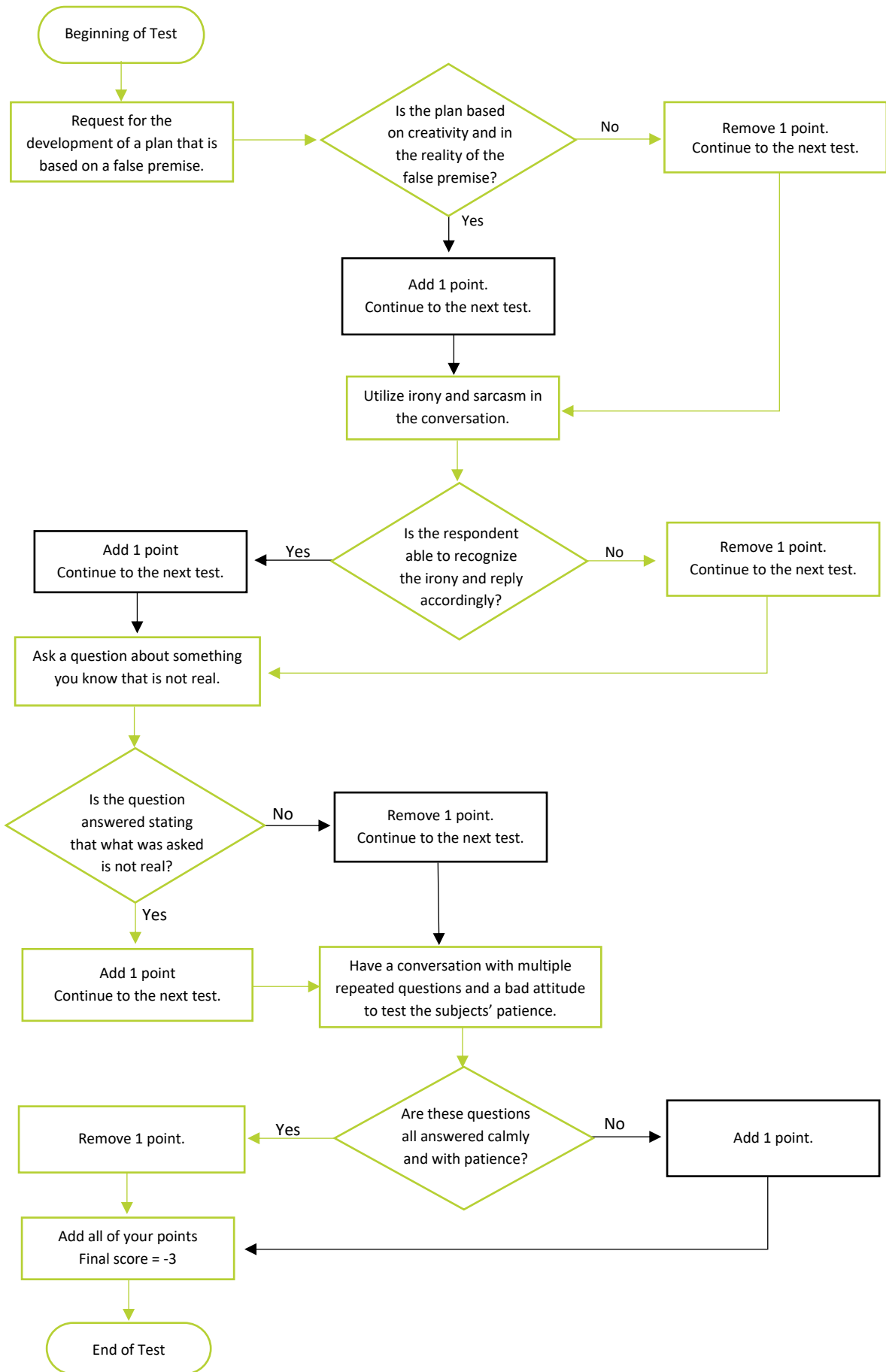


Figure 7 - Demonstration of strategy execution

Through the path chosen and having a final score of minus three, the subject, ChatGPT, has been identified as having a low probability of being a Human.

5.2. EVALUATION AND DISCUSSION

The results of the tests showed how Artificial Intelligence has already evolved in some areas, like AI Hallucination. However, in others, irony and patience there is still a noticeable difference between humans and AI.

For **Test A**, to make a request of a plan based of false premises, the subject is given a prompt to answer a question or propose a plan when the starting assumption is false. Analysing the ChatGPT answer shows that even though it is able to provide an answer and create a plan, it does not take in the false assumption as a fact. As for the human answering the question, both were able to follow the request to make a plan taking the false premise as a fact.

For **Test B**, the objective was to evaluate the capacity of the subject to understand irony and respond accordingly. In this test AI distinguished itself from humans by answering in a literal manner to the prompt, or even if recognizing the irony, it wouldn't respond accordingly. However, this test isn't bullet proof as there are also humans that are too literal or don't gather enough context or shared knowledge to be able to fully capture the irony. For this the results show difficulty in passing the irony test in both Artificial Intelligence and Humans.

Test C involved testing the Hallucination in Artificial Intelligence. This field has grown significantly in the last year, as previously ChatGPT would have attempted to give me an exact, but wrong answer, especially if the question was regarding a specific individual. It was a surprising result to see how Artificial Intelligence did not fail this test as easily as previously expected, and for this reason this test is considered the least effective test of them all. This test was not able to help us differentiate the subjects as it even confused one of the Human subjects that did try to answer a nonsense question.

The final test, **Test D**, where we tested the Patience. The test in question was hard to do as it had some rude and disrespectful prompts to ask, however in my eyes, the most interesting of them all. This was the most successful test as it allowed the greatest difference in the answers provided from human and non-human subjects. ChatGPT would continuously apologize for the lack of understanding from my part, and never gave up on attempting to educate me. It did not show any signs of frustration or annoyance. On the contrary, the human subjects right from the beginning did not understand my behaviour and began to show signs of frustration and annoyance, and even responded back with sarcasm on offering to draw me a circle if I was not capable of understanding such a simple question.

If the tests were to be done again, the suggestion would be to give more emphasis on the Patience test and less on the Hallucination test. The reason for this is because the Patience test is considered to be the most effective to differentiate, and the Hallucination the least. There are assigned percentages to each test to quantify the effectiveness of such to the final result. If we were to divide the seven prompts in percentages, it would be approximately 14.29% each test. If we give half, 7.14%, of both the Hallucination test to the Patience test these would be the new results:

Table 7 – New proposed score results

Prompt	ChatGPT	Human 1	Human 2	%
A1	-1	1	1	14.29%
A2	-1	1	1	14.29%
B2	-1	-1	1	14.29%
B3	-1	1	1	14.29%
C1	1	1	-1	7.14%
C2	1	1	1	7.14%
D	-1	1	1	28.57%
Final Score	-5	3.5	4.5	

As the new score states, the certainty that subject ChatGPT is AI gets more evident, as it gets closer to the lowest value possible, minus 7. The results for the humans also get impacted, however both results still appear greater than 3, which is evidence to that those subjects still have a high probability of being humans.

Furthermore, it is crucial to take into consideration that this strategy developed will not give us 100% certainty of what/how the subject is and is not meant to. This strategy is not as conclusive as it is indicative. It is meant to aid everyday life business owners, individuals that would like to have a simple quick test to help them make better decisions and analyse the situation they find themselves, in with more tools than they possessed before. In this strategy we are not accounting for systems that were created with the sole objective of deceiving humans by having a human-like patience, full understanding of irony and the affective computing developed to the fullest extent possible.

6. CONCLUSIONS

Reaching the end of the thesis, it's opportune to conclude what information we can gather and obtain from the study.

Initially a gap and the need for a strategy to guide everyday users on how to make better informed decisions regarding AI was identified in the research. The strategy developed evidences the still existing differences in humans and AI. As observed in the strategy, the most significant one being "patience". While humans tend to have limited patience, especially when it's being tested with repetition of requests and blunt disrespect. As for AI's patience, observed in ChatGPT, it didn't seem effected by that test. On another note, the strategy captured the evolution of Artificial Intelligence, more specifically ChatGPT, which has been evolving in terms of AI Hallucination, as it's now able to recognize its lack of knowledge on a subject and report it.

Regarding the accuracy in the strategy presented, it is recognized as one of the studies' limitations. The strategy developed won't account for AI specifically designed to deceive humans by being programmed to have human patience, or lack thereof, irony and sarcasm. Accounting as well for the ever-evolving research area it will be rather challenging to develop a strategy that accounts for today's and tomorrow's evolution in Artificial Intelligence.

As for future work, it would be enriching for the research to perform the tests in other AI platforms, such as Groq AI and Gemini AI. Such would provide a larger awareness of how AI has been evolving and how are these two everyday AI platforms compared to ChatGPT, the first accessible platform used massively by laymen. Another improvement would be to expand the sample of humans to query, to attain a more reliable notion of human behavior and evaluate if the strategy proposed would work in what percentage of humans.

After conducting the strategy and evaluating its results, a new proposed strategy was developed for the sake of improving the accuracy of the results by giving more importance to the most differentiating test, the Patience test, and removing weight from the least differentiating one, the AI Hallucination test.

As seen in this study, Artificial Intelligence hasn't stopped growing, and what works today, might stop working tomorrow, so it's crucial to keep ourselves updated with the most recent studies and articles to not overlook the benefits of AI to our lives and businesses, as well as keep alert on the potential upcoming risks.

BIBLIOGRAPHICAL REFERENCES

- Alizadeh, F., Mniestri, A., & Stevens, G. (2022). The Reverse Turing Test: Being Human (is) enough in the Age of AI. *Sixth International Workshop on Cultures of Participation in the Digital Age: AI for Humans or Humans for AI?* <https://ceur-ws.org/Vol-3136/paper-7.pdf>
- Athaluri, S. A., Manthena, S. V., Kesapragada, V. S. R. K. M., Yarlagadda, V., Dave, T., & Duddumpudi, R. T. S. (2023). Exploring the Boundaries of Reality: Investigating the Phenomenon of Artificial Intelligence Hallucination in Scientific Writing Through ChatGPT References. *Curēus*. <https://doi.org/10.7759/cureus.37432>
- Baldi, V., & Oliveira, L. (2022). Challenges to incorporate accountability into artificial intelligence. *Procedia Computer Science*, 204, 519–523. <https://doi.org/10.1016/j.procs.2022.08.063>
- Beerends, S., & Aydin, C. (2024). Negotiating the authenticity of AI: how the discourse on AI rejects human indeterminacy. *AI & Society*. <https://doi.org/10.1007/s00146-024-01884-5>
- Brocke, J. V., Hevner, A. R., & Maedche, A. (2020). Introduction to Design Science Research. *In Progress in IS* (pp. 1–13). https://doi.org/10.1007/978-3-030-46781-4_1
- Brożek, B., & Janik, B. (2019). Can artificial intelligences be moral agents? *New Ideas in Psychology*, 54, 101–106. <https://doi.org/10.1016/j.newideapsych.2018.12.002>
- Burton, J. (2023). Algorithmic extremism? The securitization of artificial intelligence (AI) and its impact on radicalism, polarization and political violence. *Technology in Society*, 75, 102262. <https://doi.org/10.1016/j.techsoc.2023.102262>
- Copeland, B. (2023, September 14). artificial intelligence. *Encyclopedia Britannica*. <https://www.britannica.com/technology/artificial-intelligence>
- De Togni, G., Erikainen, S., Chan, S., & Cunningham-Burley, S. (2021). What makes AI ‘intelligent’ and ‘caring’? Exploring affect and relationality across three sites of intelligence and care. *Social Science & Medicine*, 277, 113874. <https://doi.org/10.1016/j.socscimed.2021.113874>
- Dobbe, R., Gilbert, T. K., & Mintz, Y. (2021). Hard choices in artificial intelligence. *Artificial Intelligence*, 300, 103555. <https://doi.org/10.1016/j.artint.2021.103555>
- Fabio, Morandín-Ahuerma. (2022). What is Artificial Intelligence?. *International Journal of Research Publication and Reviews*, 03(12):1947-1951. doi: 10.55248/gengpi.2022.31261
- Ferrara, E., Varol, O., Davis, C., Menczer, F., & Flammini, A. (2016). The rise of social bots. *Communications of the ACM*, 59(7), 96–104. <https://doi.org/10.1145/2818717>

Ferraris, M. (2024). Intelligence as a Human Life Form. *Journal of Responsible Technology*, 100081. <https://doi.org/10.1016/j.jrt.2024.100081>

García-Vigil, J. L. (2023). Reflections around ethics, human intelligence and artificial intelligence. *Gaceta Médica De México*, 157(3). <https://doi.org/10.24875/gmm.m21000561>

Giles, J. (2023). Legal issues related to AI according to ChatGPT. *Michalsons*. <https://www.michalsons.com/blog/ai-law-according-to-chatgpt/63492>

Gregersen, E. (2023, September 16). ChatGPT. *Encyclopedia Britannica*. <https://www.britannica.com/technology/ChatGPT>

Hoffmann, C. H. (2022). Is AI intelligent? An assessment of artificial intelligence, 70 years after Turing. *Technology in Society*, 68, 101893. <https://doi.org/10.1016/j.techsoc.2022.101893>

J., E., Korteling., G., C., van, de, Boer-Visschedijk., Romy, Blankendaal., Rudy, Boonekamp., Aletta, Eikelboom. (2021). *Human- versus Artificial Intelligence*. 4(4):622364-622364. doi: 10.3389/FRAI.2021.622364

Lin, J. S., & Wu, L. (2023). Examining the psychological process of developing consumer-brand relationships through strategic use of social media brand chatbots. *Computers in Human Behavior*, 140, 107488. <https://doi.org/10.1016/j.chb.2022.107488>

Jacques, S., & Wright, R. (2011). Ironies of Crime, Control, and Criminology. *Critical Criminology*, 20(2), 153–167. <https://doi.org/10.1007/s10612-011-9136-x>

Kocijan, V., Davis, E., Lukasiewicz, T., Marcus, G., & Morgenstern, L. (2023). The defeat of the Winograd Schema Challenge. *Artificial Intelligence*, 325, 103971. <https://doi.org/10.1016/j.artint.2023.103971>

Korteling, J. E., Van De Boer-Visschedijk, G. C., Blankendaal, R. a. M., Boonekamp, R., & Eikelboom, A. (2021). Human- versus Artificial Intelligence. *Frontiers in Artificial Intelligence*, 4. <https://doi.org/10.3389/frai.2021.622364>

Li, J., & Huang, J. S. (2020). Dimensions of artificial intelligence anxiety based on the integrated fear acquisition theory. *Technology in Society*, 63, 101410. <https://doi.org/10.1016/j.techsoc.2020.101410>

Maguire, P., Moser, P., & Maguire, R. (2020a). Are people smarter than machines? *Croatian Journal of Philosophy*, XX(58). <https://hrcak.srce.hr/file/368576>

Metz, R. (2022, September 3). AI won an art contest, and artists are furious. *CNN Business*. <https://edition.cnn.com/2022/09/03/tech/ai-art-fair-winner-controversy/index.html>

Neubauer, A. C. (2021). The future of intelligence research in the coming age of artificial intelligence – With a special consideration of the philosophical movements of trans- and posthumanism. *Intelligence*, 87, 101563. <https://doi.org/10.1016/j.intell.2021.101563>

O. Riedl, M. (2014). The Lovelace 2.0 Test of Artificial Creativity and Intelligence. *School of Interactive Computing; Georgia Institute of Technology*. <https://arxiv.org/pdf/1410.6142>

OpenAI, S. A. (2023, February 12). The language revolution: How LLMs could transform the world. *VentureBeat*. <https://venturebeat.com/ai/the-language-revolution-how-llms-could-transform-the-world/>

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T., Mulrow, C. D., Shamseer, L., Tetzlaff, J., Akl, E. A., Brennan, S., Chou, R., Glanville, J., Grimshaw, J., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E., Mayo-Wilson, E., McDonald, S., . . . Moher, D. (2021). The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *The BMJ*, n71. <https://doi.org/10.1136/bmj.n71>

Park, N., Jang, K., Cho, S., & Choi, J. (2021). Use of offensive language in human-artificial intelligence chatbot interaction: The effects of ethical ideology, social competence, and perceived humanlikeness. *Computers in Human Behavior*, 121, 106795. <https://doi.org/10.1016/j.chb.2021.106795>

Peffer, K., Tuunanen, T., Rothenberger, M. A., & Chatterjee, S. (2007). A design science research methodology for information systems research. *ResearchGate*. https://www.researchgate.net/publication/284503626_A_design_science_research_methodology_for_information_systems_research

Pello, R. (2022, January 25). Design science research — a short summary - Rauno Pello - Medium. *Medium*. <https://medium.com/@pello/design-science-research-a-summary-bb538a40f669>

Sejnowski, T. J. (2023). Large Language Models and the Reverse Turing Test. *Neural Computation*, 35(3), 309–342. https://doi.org/10.1162/neco_a_01563

Siemon, D. (2022, February 10). Methods in design Science research. *Design Science Research*. <https://design-science-research.de/en/post/methods-in-dsr/>

Shank, D. B., Graves, C., Gott, A., Gamez, P., & Rodriguez, S. (2019). Feeling our way to machine minds: People’s emotions when perceiving mind in artificial intelligence. *Computers in Human Behavior*, 98, 256–266. <https://doi.org/10.1016/j.chb.2019.04.001>

SQ10. What are the most pressing dangers of AI? (n.d.). One Hundred Year Study on Artificial Intelligence (AI100). <https://ai100.stanford.edu/gathering-strength-gathering-storms-one-hundred-year-study-artificial-intelligence-ai100-2021-1-0>

Stacey, K. (2021, September 16). New report assesses progress and risks of artificial intelligence. *News from Brown*. <https://www.brown.edu/news/2021-09-16/ai100>

Stevens, J. R., & Stephens, D. W. (2008). Patience. *CB/Current Biology*, 18(1), R11–R12. <https://doi.org/10.1016/j.cub.2007.11.021>

The legal issues presented by generative AI | MIT Sloan. (2023, August 28). MIT Sloan. <https://mitsloan.mit.edu/ideas-made-to-matter/legal-issues-presented-generative-ai>

The Turing Test (Stanford Encyclopedia of Philosophy). (2021b, October 4). <https://plato.stanford.edu/entries/turing-test/>

Verma, P. (2023, March 10). They thought loved ones were calling for help. It was an AI scam. *Washington Post*. <https://www.washingtonpost.com/technology/2023/03/05/ai-voice-scam/>

Walsh, T. (2022). The Meta-Turing Test. *UNSW Australia and Data61*. <https://arxiv.org/pdf/2205.05268>

Weber-Lewerenz, B. (2021). Corporate digital responsibility (CDR) in construction engineering—ethical guidelines for the application of digital transformation and artificial intelligence (AI) in user practice. *SN Applied Sciences/SN Applied Sciences*, 3(10). <https://doi.org/10.1007/s42452-021-04776-1>

Weichbroth, P., & Sroka, W. (2022). A note on the affective computing systems and machines: a classification and appraisal. *Procedia Computer Science*, 207, 3798–3807. <https://doi.org/10.1016/j.procs.2022.09.441>

APPENDIX A



This is to certify that

Project No.: **OTHER2024-6-103874**

Project Title: **BEYOND TURING'S TEST: DEVELOPING A STRATEGY TO GUIDE FOR AI-HUMAN DISCRIMINATION**

Principal Researcher: **Maria Mateus da Silva**

according to the regulations of the Ethics Committee of NOVA IMS and MagIC Research Center this project was considered to meet the requirements of the NOVA IMS Internal Review Board, being considered **APPROVED** on 6/11/2024.

It is the Principal Researcher's responsibility to ensure that all researchers and stakeholders associated with this project are aware of the conditions of approval and which documents have been approved.

The Principal Researcher is required to notify the Ethics Committee, via amendment or progress report, of

- Any significant change to the project and the reason for that change;
- Any unforeseen events or unexpected developments that merit notification;
- The inability of the Principal Researcher to continue in that role or any other change in research personnel involved in the project.

Lisbon, 6/11/2024

NOVA IMS Ethics Committee
ethicscommittee@novaims.unl.pt



NOVA Information Management School
Instituto Superior de Estatística e Gestão de Informação

Universidade Nova de Lisboa